In [2]:

```
# Load the Drive helper and mount
from google.colab import drive

# This will prompt for authorization.
drive.mount('/content/drive')
```

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6
qk8qdgf4n4g3pfee6491hc0brc4i.apps.googleusercontent.com&redirect_uri=urn%3Aietf%3Awg%3Aoauth%3A2.0%
b&scope=email%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdocs.test%20https%3A%2F%2Fwww.googleapis.
2Fauth%2Fdrive%20https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdrive.photos.readonly%20https%3A%2F%2Fww
ogleapis.com%2Fauth%2Fpeopleapi.readonly&response_type=code

Enter your authorization code:
··········
Mounted at /content/drive
```

Cluster

In [35]:

```
import pandas as pd
import numpy as np

df = pd.read_csv('drive/My Drive/Colab Notebooks/cluster/review.csv')
df.head(9)
```

Out[35]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summar
0	1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Goo Quali Dog Foo
1	2	B00813GRG4	A1D87F6ZCVE5NK	dll pa	0	0	1	1346976000	Not a Advertise
2	3	B000LQOCH0	ABXLMWJIXXAIN	Natalia Corres "Natalia Corres"	1	1	4	1219017600	"Deligh says it a
3	4	B000UA0QIQ	A395BORC6FGVXV	Karl	3	3	2	1307923200	Coug Medicir
4	5	B006K2ZZ7K	A1UQRSCLF8GW1T	Michael D. Bigham "M. Wassir"	0	0	5	1350777600	Great taf
5	6	B006K2ZZ7K	ADT0SRK1MGOEU	Twoapennything	0	0	4	1342051200	Nice Taf
									Great! Ju as goo

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary
6		B006K2ZZ7K	A1SP2KVKFXXRU1	David C. Sullivan	0	0	5	1340150400	expensi brand
7	8	B006K2ZZ7K	A3JRGQVEQN31IQ	Pamela G. Williams	0	0	5	1336003200	Wonderfu tasty taf
8	9	B000E7L2R4	A1MZYO9TZK0BBI	R. James	1	1	5	1322006400	Yay Barle

In [2]:

```
# Load the Drive helper and mount
from google.colab import drive

# This will prompt for authorization.
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
```

Pre-processed the given data

In [1]:

```
import pickle
final = pickle.load(open('drive/My Drive/Colab Notebooks/cluster/final.p','rb'))

# ensuring time based split
from sklearn.model_selection import train_test_split
##Sorting data according to Time in ascending order for Time Based Splitting
time_sorted_data = final.sort_values('Time', axis=0, ascending=True, inplace=False, kind='quicksort
', na_position='last')
final.head(2)
```

Out[1]:

	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summ
138706	150524	0006641040	ACITT7DI6IDDL	shari zychinski	0	0	1	939340800	EV bo educati
138683	150501	0006641040	AJ46FKXOVC7NR	Nicholas A Mesiano	2	2	1	940809600	This w seri great to sp time

In [30]:

```
%env JOBLIB_TEMP_FOLDER=/tmp
```

```
env: JOBLIB_TEMP_FOLDER=/tmp
```

In [5]:

```
!total cpu : !lscpu
```

```
/bin/bash: total: command not found
```

In [0]:

```
```

In [2]:

```
# sampling the data for ease

x = final['CleanedText'].iloc[:100000]
x.shape
```

Out[2]:

```
(100000,)
```

Avg W2V

In [0]:

```
#IMporting files from minbatch
avg = pickle.load(open('avg1.p','rb'))
```

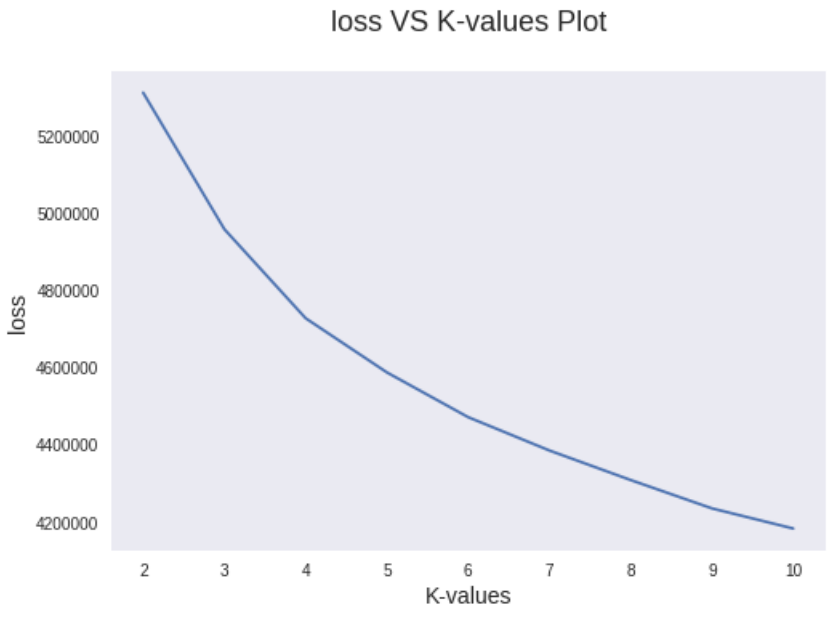
In [0]:

```
from sklearn.cluster import KMeans

k_values = [2,3,4,5,6,7,8,9,10]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(avg)
    loss.append(kmeans.inertia_)
```

In [7]:

```
# Loss plot
import matplotlib.pyplot as plt
# Draw Loss VS K values plot
plt.plot(k_values, loss)
plt.xlabel('K-values',size=14)
plt.ylabel('loss',size=14)
plt.title('loss VS K-values Plot\n',size=18)
plt.grid()
plt.show()
```



In [0]:

```
from sklearn.cluster import MiniBatchKMeans
```

```
# speculating rate of chage from the plot is at 6
optimal_k = 6

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(avg)
```

In [15]:

```
# review clusters

reviews = df['Text'].values
# Getting all the reviews in different clusters
cluster0 = []
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
cluster5 = []

check=[]


for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster0.append(reviews[i])

    elif kmeans.labels_[i] == 1:
        cluster1.append(reviews[i])

    elif kmeans.labels_[i] == 2:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 3:
        cluster3.append(reviews[i])
    elif kmeans.labels_[i] == 4:
        cluster4.append(reviews[i])
    elif kmeans.labels_[i] == 5:
        cluster5.append(reviews[i])

    else:
        check.append(reviews[i])




# Number of reviews in different clusters
print("\nNo. of reviews in Cluster-0 : ",len(cluster0))
print("\nNo. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
print("\nNo. of reviews in Cluster-5 : ",len(cluster5))

print("\nNo. of reviews in check: ",len(check))
```

```
No. of reviews in Cluster-0 :  46878

No. of reviews in Cluster-1 :  45574

No. of reviews in Cluster-2 :  66524

No. of reviews in Cluster-3 :  40732

No. of reviews in Cluster-4 :  28273

No. of reviews in Cluster-5 :  72019

No. of reviews in check:  0
```

In [17]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster0[i]))
    count +=1
```

```
    count +=1
```

```
Review-1 :
 The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be around for a while wi
th my son and I.

Review-2 :
 My daughter loves twizzlers and this shipment of six pounds really hit the spot. It's exactly wha
t you would expect...six packages of strawberry twizzlers.

Review-3 :
 This is the same stuff you can buy at the big box stores.  There is nothing healthy about it.  It
is just carbs and sugars.  Save your money and get something that at least has some taste.
```

In [18]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster1[i]))
    count +=1
```

```
Review-1 :
 I got this for my Mum who is not diabetic but needs to watch her sugar intake, and my father who
simply chooses to limit unnecessary sugar intake - she's the one with the sweet tooth - they both
LOVED these toffees, you would never guess that they're sugar-free and it's so great that you can
eat them pretty much guilt free!  i was so impressed that i've ordered some for myself (w dark cho
colate) to take to the office so i'll eat them instead of snacking on sugary sweets.<br />These ar
e just EXCELLENT!

Review-2 :
 The variety packs taste great!<br /><br />I have them every morning. At $0.30 cents per meal, I d
on't understand why everyone on earth isn't buying this stuff up.<br /><br />Maple and brown sugar
is terrific, followed by apples and cinnamon, followed by regular. You don't get tired of the same
ole thing, and they taste great.<br /><br />I just boil water from a small pot, empty the packet o
r 2 in a bowl, pour in boiling water, and watch it expand to 2x its size!<br /><br />Taste really
good and takes minutes to prepare.<br /><br />Not sure why everyone on earth isn't this.
Convenient, healthy, very quick, excellent quality, and extremely cheap...

Review-3 :
 I really like the Maple and Brown Sugar flavor. The regular is fine with brown sugar added. The A
pples and Cinnamon flavor is OK. This is a very quick, easy and satisfying breakfast and I'll orde
r this brand again, but not the variety. I'll get all Maple and Brown Sugar.
```

In [19]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster2[i]))
    count +=1
```

```
Review-1 :
 It is okay.  I would not go out of my way to buy it again

Review-2 :
 These taste really good. I have been purchasing a different brand and these are very similar in t
aste and texture. I agree with the other reviewer regarding ordering in the summer. There is no in
sulating packaging with ice packs so they will melt in warm weather like all chocolate food items.
Order in cold weather and buy enough to last!!!

Review-3 :
 My dog has a ton of allergies both environmental and food. She was on a prescription dog food bef
ore we had her tested to see what allergies she has. After we got the test back, we learned she wa
s allergic to something in the prescription brand. So I finally found this dog food and she has do
ne so well on this! She still has her environmental triggers, but I am happy she can finally eat s
omething I know won't cause her pain.
```

In [20]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster3[i]))
    count +=1
```

```
Review-1 :
 I purchased the Mango flavor, and to me it doesn't take like Mango at all.  There is no hint of sweetness, and unfortunately there is a hint or aftertaste almost like licorice.  I've been consuming various sports nutrition products for decades, so I'm familiar and have come to like the taste of the most of the products I've tried.  The mango flavor is one of the least appealing I've tasted.  It's not terrible, but it's bad enough that I notice the bad taste every sip I take.

Review-2 :
 Always being a fan of ramen as a quick and easy meal, finding it on amazon for a decent price and having it delivered to your door by the case is an amazing situation for anyone to find themselves in.

Review-3 :
 Got these on sale for roughly 25 cents per cup, which is half the price of my local grocery stores, plus they rarely stock the spicy flavors. These things are a GREAT snack for my office whe re time is constantly crunched and sometimes you can't escape for a real meal. This is one of my f avorite flavors of Instant Lunch and will be back to buy every time it goes on sale.
```

In [21]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster4[i]))
    count +=1
```

```
Review-1 :
 this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault

Review-2 :
 Tastes great, and gives me energy without the jitters. Love this product.<br />However, I do wish it came in some sort of resealable jar or hard plastic container instead of a zip-lock pouch because the zip-lock ripped off and quit working after awhile. But the product is fantastic, just not crazy about the packaging.

Review-3 :
 If used on a daily basis you will keep your gums healty and between your teeth clean and free of plack.
```

In [22]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster5[i]))
    count +=1
```

```
Review-1 :
 I have bought several of the Vitality canned dog food products and have found them all to be of g ood quality. The product looks more like a stew than a processed meat and it smells better. My Lab rador is finicky and she appreciates this product better than  most.

Review-2 :
 Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

Review-3 :
 This is a confection that has been around a few centuries.  It is a light, pillowy citrus gelatin with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with powdered sugar.  And it is a tiny mouthful of heaven.  Not too chewy, and very flavorful.  I highl y recommend this yummy treat.  If you are familiar with the story of C.S. Lewis' "The Lion, The Wi tch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Si sters to the Witch.
```

TFIDF-W2V

In [0]:

```
import pickle
final = pickle.load(open('drive/My Drive/Colab Notebooks/cluster/final.p','rb'))

# ensuring time based split
from sklearn.model_selection import train_test_split
##Sorting data according to Time in ascending order for Time Based Splitting
time_sorted_data = final.sort_values('Time', axis=0, ascending=True, inplace=False, kind='quicksort
', na_position='last')
```

In [20]:

```
X = final['CleanedText'].iloc[:300000]
X.head(2)
```

Out[20]:

```
138706    witti littl book make son laugh loud recit car...
138683    rememb see show air televis year ago child sis...
Name: CleanedText, dtype: object
```

In [21]:

```
!pip install gensim
```

```
Requirement already satisfied: gensim in /usr/local/lib/python3.6/dist-packages (3.6.0)
Requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.6/dist-packages (from gensim)
(1.11.0)
Requirement already satisfied: scipy>=0.18.1 in /usr/local/lib/python3.6/dist-packages (from
gensim) (0.19.1)
Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.6/dist-packages (from
gensim) (1.7.1)
Requirement already satisfied: numpy>=1.11.3 in /usr/local/lib/python3.6/dist-packages (from
gensim) (1.14.6)
Requirement already satisfied: bz2file in /usr/local/lib/python3.6/dist-packages (from smart-
open>=1.2.1->gensim) (0.98)
Requirement already satisfied: boto3 in /usr/local/lib/python3.6/dist-packages (from smart-
open>=1.2.1->gensim) (1.9.38)
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packages (from smart-
open>=1.2.1->gensim) (2.18.4)
Requirement already satisfied: boto>=2.32 in /usr/local/lib/python3.6/dist-packages (from smart-
open>=1.2.1->gensim) (2.49.0)
Requirement already satisfied: jmespath<1.0.0,>=0.7.1 in /usr/local/lib/python3.6/dist-packages
(from boto3->smart-open>=1.2.1->gensim) (0.9.3)
Requirement already satisfied: s3transfer<0.2.0,>=0.1.10 in /usr/local/lib/python3.6/dist-packages
(from boto3->smart-open>=1.2.1->gensim) (0.1.13)
Requirement already satisfied: botocore<1.13.0,>=1.12.38 in /usr/local/lib/python3.6/dist-packages
(from boto3->smart-open>=1.2.1->gensim) (1.12.38)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from
requests->smart-open>=1.2.1->gensim) (2018.10.15)
Requirement already satisfied: urllib3<1.23,>=1.21.1 in /usr/local/lib/python3.6/dist-packages
(from requests->smart-open>=1.2.1->gensim) (1.22)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages
(from requests->smart-open>=1.2.1->gensim) (3.0.4)
Requirement already satisfied: idna<2.7,>=2.5 in /usr/local/lib/python3.6/dist-packages (from
requests->smart-open>=1.2.1->gensim) (2.6)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1; python_version >= "2.7" in
/usr/local/lib/python3.6/dist-packages (from botocore<1.13.0,>=1.12.38->boto3->smart-open>=1.2.1->
gensim) (2.5.3)
Requirement already satisfied: docutils>=0.10 in /usr/local/lib/python3.6/dist-packages (from
botocore<1.13.0,>=1.12.38->boto3->smart-open>=1.2.1->gensim) (0.14)
```

In [0]:

```
sent_of_sent=[]
for sent in X:
    sent_of_sent.append(sent.split())
```

In [23]:

```
#word to vector
from gensim.models import Word2Vec
w2v_model=Word2Vec(sent_of_sent,min_count=3,size=200, workers=4)# words which occurs 3 times; 500 dimensions

w2v_words = list(w2v_model.wv.vocab)
print("number of words that occured minimum 3 times ",len(w2v_words))
```

```
number of words that occured minimum 3 times  26564
```

In [0]:

```

```

In [0]:

```
sent_of_sent=[]
for sent in X:
    sent_of_sent.append(sent.split())
```

In [0]:

```
from sklearn.feature_extraction.text import TfidfVectorizer

m = TfidfVectorizer()
tf_idf_matrix = m.fit_transform(X)
# we are converting a dictionary with word as a key, and the idf as a value
dictionary = dict(zip(m.get_feature_names(), list(m.idf_)))
```

In [26]:

```
from tqdm import tqdm
import numpy as np

# TF-IDF weighted Word2Vec
tfidf_feat = m.get_feature_names() # tfidf words/col-names
# final_tf_idf is the sparse matrix with row= sentence, col=word and cell_val = tfidf

tfidf_sent_vectors = []; # the tfidf-w2v for each sentence/review is stored in this list
row=0;
for sent in tqdm(sent_of_sent): # for each review/sentence
    sent_vec = np.zeros(200) # as word vectors are of zero length
    weight_sum =0; # num of words with a valid vector in the sentence/review
    for word in sent: # for each word in a review/sentence
        if word in w2v_words:
            vec = w2v_model.wv[word]
            tf_idf = dictionary[word]*(sent.count(word)/len(sent))
            sent_vec += (vec * tf_idf)
            weight_sum += tf_idf
    if weight_sum != 0:
        sent_vec /= weight_sum
    tfidf_sent_vectors.append(sent_vec)
    row += 1
```

```
100%|██████████| 300000/300000 [14:27<00:00, 345.87it/s]
```

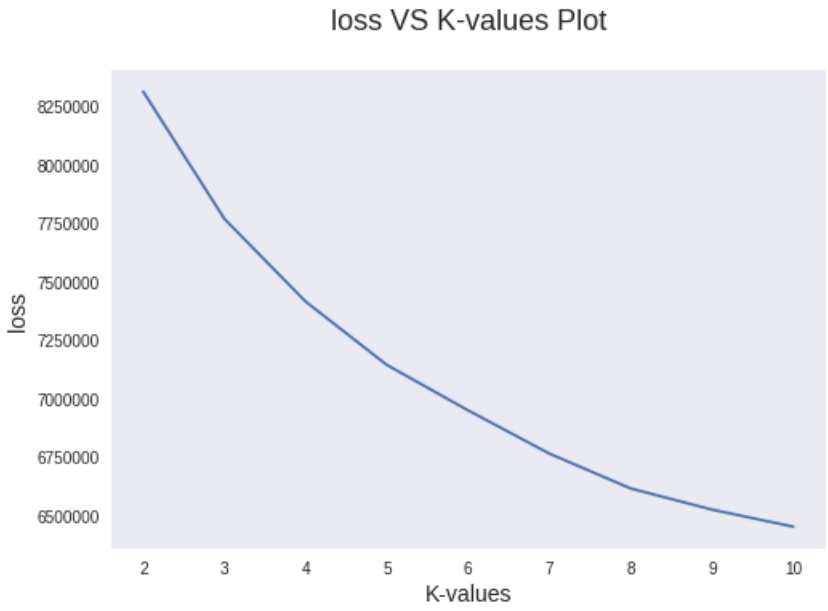
In [0]:

```
from sklearn.cluster import KMeans

k_values = [2,3,4,5,6,7,8,9,10]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(tfidf_sent_vectors)
    loss.append(kmeans.inertia_)
```

In [32]:

```
import matplotlib.pyplot as plt
# Draw Loss VS K values plot
plt.plot(k_values, loss)
plt.xlabel('K-values',size=14)
plt.ylabel('loss',size=14)
plt.title('loss VS K-values Plot\n',size=18)
plt.grid()
plt.show()
```



In [0]:

```
# speculating rate of chage from the plot is at 8
optimal_k = 8

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(tfidf_sent_vectors)
```

In [39]:

```
reviews = df['Text'].values
# Getting all the reviews in different clusters
cluster0 = []
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
cluster5 = []
cluster6 = []
cluster7 = []
check=[]


for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster0.append(reviews[i])

    elif kmeans.labels_[i] == 1:
        cluster1.append(reviews[i])

    elif kmeans.labels_[i] == 2:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 3:
        cluster3.append(reviews[i])
    elif kmeans.labels_[i] == 4:
        cluster4.append(reviews[i])
    elif kmeans.labels_[i] == 5:
        cluster5.append(reviews[i])
    elif kmeans.labels_[i] == 6:
        cluster6.append(reviews[i])
```

```
    elif kmeans.labels_[i] == 7:
        cluster7.append(reviews[i])

    else:
        check.append(reviews[i])



# Number of reviews in different clusters
print("\nNo. of reviews in Cluster-0 : ",len(cluster0))
print("\nNo. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
print("\nNo. of reviews in Cluster-5 : ",len(cluster5))
print("\nNo. of reviews in Cluster-6 : ",len(cluster6))
print("\nNo. of reviews in Cluster-7 : ",len(cluster7))

print("\nNo. of reviews in check: ",len(check))
```

```
No. of reviews in Cluster-0 :  40134

No. of reviews in Cluster-1 :  24575

No. of reviews in Cluster-2 :  42046

No. of reviews in Cluster-3 :  71803

No. of reviews in Cluster-4 :  43488

No. of reviews in Cluster-5 :  20739

No. of reviews in Cluster-6 :  13836

No. of reviews in Cluster-7 :  43379

No. of reviews in check:  0
```

In [40]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster0[i]))
    count +=1
```

```
Review-1 :
 I purchased the Mango flavor, and to me it doesn't take like Mango at all.  There is no hint of sweetness, and unfortunately there is a hint or aftertaste almost like licorice.  I've been consuming various sports nutrition products for decades, so I'm familiar and have come to like the taste of the most of the products I've tried.  The mango flavor is one of the least appealing I've tasted.  It's not terrible, but it's bad enough that I notice the bad taste every sip I take.

Review-2 :
 Always being a fan of ramen as a quick and easy meal, finding it on amazon for a decent price and having it delivered to your door by the case is an amazing situation for anyone to find themselves in.

Review-3 :
 Got these on sale for roughly 25 cents per cup, which is half the price of my local grocery stores, plus they rarely stock the spicy flavors. These things are a GREAT snack for my office whe re time is constantly crunched and sometimes you can't escape for a real meal. This is one of my f avorite flavors of Instant Lunch and will be back to buy every time it goes on sale.
```

In [41]:

```
# Three Reviews of cluster 1
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster1[i]))
    count +=1
```

```
Review-1 :
 this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault

Review-2 :
 Natural Balance Dry Dog Food Lamb Meal and Brown Rice Recipe did wonders for my Jack Russell. She
has awful food allergy's and this food was our last hope because it was the last food we could fin
d that didn't have SOMETHING in it that she was allergic to. She has no problem eating it dry but
normally I mix the Natural Balance Dry Lamb and Brown Rice with the Natural Balance Wet Lamb and B
rown Rice.. she seems to like that better. We started feeding it to our other dog (a Bichon) too a
nd she loves it. If your dog has allergy's or stomach issue.. or if you want your dog to eat bette
r food-- this is it! You will see a difference in your pet.

Review-3 :
 Tastes great, and gives me energy without the jitters. Love this product.<br />However, I do wish
it came in some sort of resealable jar or hard plastic container instead of a zip-lock pouch
because the zip-lock ripped off and quit working after awhile. But the product is fantastic, just
not crazy about the packaging.
```

In [42]:

```
# Three Reviews of cluster 2
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster2[i]))
    count +=1
```

```
Review-1 :
 The Strawberry Twizzlers are my guilty pleasure - yummy. Six pounds will be around for a while wi
th my son and I.

Review-2 :
 Product received is as advertised.<br /><br /><a
href="http://www.amazon.com/gp/product/B001GVISJM">Twizzlers, Strawberry, 16-Ounce Bags (Pack of 6
)</a>

Review-3 :
 Good oatmeal.  I like the apple cinnamon the best.  Though I wouldn't follow the directions on th
e package since it always comes out too soupy for my taste.  That could just be me since I like my
oatmeal really thick to add some milk on top of.
```

In [43]:

```
# Three Reviews of cluster 3
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster3[i]))
    count +=1
```

```
Review-1 :
 I have bought several of the Vitality canned dog food products and have found them all to be of g
ood quality. The product looks more like a stew than a processed meat and it smells better. My Lab
rador is finicky and she appreciates this product better than  most.

Review-2 :
 Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted.
Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".

Review-3 :
 This is a confection that has been around a few centuries.  It is a light, pillowy citrus gelatin
with nuts - in this case Filberts. And it is cut into tiny squares and then liberally coated with
powdered sugar.  And it is a tiny mouthful of heaven.  Not too chewy, and very flavorful.  I highl
y recommend this yummy treat.  If you are familiar with the story of C.S. Lewis' "The Lion, The Wi
tch, and The Wardrobe" - this is the treat that seduces Edmund into selling out his Brother and Si
sters to the Witch.
```

In [45]:

```
# Three Reviews of cluster 4
count=1
```

```
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster4[i]))
    count +=1
```

```
Review-1 :
 Grape gummy bears are hard to find in my area. In fact pretty much anyone I talk to about grape
gummy bears they think I'm lying. So I bought 10lbs... : ) These bears are a little bit bigger the
n the other brands and have kind of sour kick, but nothing to strong. I love grape flavored
candy/soda and these are pretty good. There is another company that makes grape gummy bears that
are a little bit better in my opinion, but these are well worth it for the price. I like to use th
e gummy bears in home made Popsicles with flavored sports drink. The salt in the sports drink make
s for softer popsicles, and the gummy bears are awesome frozen. They are delicious!

Review-2 :
 The taste was great, but the berries had melted.  May order again in winter. If you order in cold
weather you should enjoy flavor.

Review-3 :
 Great gift for all ages! I purchased these giant canes before and the recipients loved them so mu
ch, they kept them and would not eat them.
```

In [44]:

```
# Three Reviews of cluster 5
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster5[i]))
    count +=1
```

```
Review-1 :
 I got this for my Mum who is not diabetic but needs to watch her sugar intake, and my father who
simply chooses to limit unnecessary sugar intake - she's the one with the sweet tooth - they both
LOVED these toffees, you would never guess that they're sugar-free and it's so great that you can
eat them pretty much guilt free!  i was so impressed that i've ordered some for myself (w dark cho
colate) to take to the office so i'll eat them instead of snacking on sugary sweets.<br />These ar
e just EXCELLENT!

Review-2 :
 The variety packs taste great!<br /><br />I have them every morning. At $0.30 cents per meal, I d
on't understand why everyone on earth isn't buying this stuff up.<br /><br />Maple and brown sugar
is terrific, followed by apples and cinnamon, followed by regular. You don't get tired of the same
ole thing, and they taste great.<br /><br />I just boil water from a small pot, empty the packet o
r 2 in a bowl, pour in boiling water, and watch it expand to 2x its size!<br /><br />Taste really
good and takes minutes to prepare.<br /><br />Not sure why everyone on earth isn't this.
Convenient, healthy, very quick, excellent quality, and extremely cheap...

Review-3 :
 I really like the Maple and Brown Sugar flavor. The regular is fine with brown sugar added. The A
pples and Cinnamon flavor is OK. This is a very quick, easy and satisfying breakfast and I'll orde
r this brand again, but not the variety. I'll get all Maple and Brown Sugar.
```

In [46]:

```
# Three Reviews of cluster 6
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster6[i]))
    count +=1
```

```
Review-1 :
 It is chocolate, what can I say.  Great variety of everything our family loves.  With a family of
six it goes fast here.  Perfect variety.  Kit Kat, Reeses, take five and more.

Review-2 :
 The mouth says, "How do I love thee, let me count the ways..."<br />If you like apple products a
must have item.  The only draw back, shipping cost.  These are very heavy.

Review-3 :
 The crust on these tarts are perfect.  My husband loves these, but I'm not so crazy about them.
They are just too sour/tart for my taste.  I'll eat the crust and hubby takes my filling.  My kids
think they're great, so maybe it's just me.
```

In [47]:

```
# Three Reviews of cluster 7
count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster7[i]))
    count +=1
```

```
Review-1 :
 this has to be one of the best teas I have ever tasted... it's clean, bright, fresh...<br /><br />
>great delivery...again quality... just try it...

Review-2 :
 This cat food was recommended by my vet because my 13 year old Cleo Cat was having trouble keepin
g any food down and was loosing weight.  It stopped her problems.  She doesn't eat as much as befo
re since she seems to be more satisfied after eating which helps justify the added price.  It aids
in digestion since the...ok sorry...smells are gone if ya know what I mean.  Had to add that so an
yone experiencing problems with their cats can see there are a lot of benifits from this cat food.

Review-3 :
 I bought this coffee because its much cheaper than the ganocafe and has the organic reishi mushro
om as well as other healthy antioxidants.  I didn't expect it to taste good, but it actually does!
I've only had it for a few days and for $5 its totally worth it.  My sisters all take ganocafe but
now I'm introducing them to this less expensive similar coffee.  I will follow up on this product
in a few weeks.  :)
```

In [49]:

```
df = pd.read_csv('drive/My Drive/Colab Notebooks/cluster/jxe69-pyth5 - jxe69-pyth5.csv')
df
```

Out[49]:

	Model	Vectoriser	Number of clusters
0	Minbatch	BOW	6
1	Minbatch	TFIDF	6
2	Minbatch	AVG-W2V	5
3	Minbatch	TFID-W2V	9
4	********	*********	**********
5	Kmeans	AVG-W2V	6
6	Kmeans	TFIDF-W2V	8